GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 10:44:53 ; Search time 58 Seconds
(without alignments)
798.927 Million cell updates/sec

Title: US-10-020-618-1
Perfect score: 852
Sequence: 1 MASPHQEPKPGDLIEIFRLG.....ILVVAGCSFXIRRYQKKATA 164

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	851	99.9	164	2	AAW85736	Aaw85736 Human ana
2	850	99.8	164	2	AAW29736	Aaw29736 Human H-r
3	840	98.6	164	2	AAW60726	Aaw60726 Full leng
4	840	98.6	164	4	AAE12165	Aae12165 Human taz
5	839	98.5	164	2	AAW60727	Aaw60727 Tazaroten
6	839	98.5	164	4	AAE12166	Aae12166 Human taz
7	839	98.5	164	6	ABR58620	Abr58620 Human can
8	839	98.5	164	6	ABR92144	Abr92144 Human cer
9	827	97.1	163	2	AAY12385	Aay12385 Human 5'
10	623	73.1	124	2	AAW60725	Aaw60725 Consensus
11	623	73.1	124	4	AAE12164	Aae12164 Protein c
12	423.5	49.7	160	7	ADD47497	Add47497 Rat Prote
13	423.5	49.7	161	7	ADD48283	Add48283 Rat Prote
14	418.5	49.1	162	4	AAM39727	Aam39727 Human pol
15	418.5	49.1	162	7	ADC37331	Adc37331 Nuclear f
16	418.5	49.1	162	7	ADD48285	Add48285 Human Pro
17	418.5	49.1	169	4	AAM41513	Aam41513 Human pol
18	417.5	49.0	162	7	ADD47499	Add47499 Human Pro
19	415.5	48.8	169	3	AAB56763	Aab56763 Human pro
20	411.5	48.3	162	7	ADC37329	Adc37329 Nuclear f
21	350	41.1	168	2	AAW85735	Aaw85735 Human ana
22	350	41.1	168	4	AAM51616	Aam51616 Human A-C
23	350	41.1	168	4	AAB60468	Aab60468 Human cel
24	345	40.5	237	4	AAM95184	Aam95184 Human rep
25	345	40.5	237	4	ABB95884	Abb95884 Human tes
26	332.5	39.0	167	4	AAM51615	Aam51615 Murine A-
27	258	30.3	108	3	AAG01863	Aag01863 Human sec
28	165.5	19.4	156	4	AAO12134	Aao12134 Human pol
29	136	16.0	310	6	AAE36006	Aae36006 Human BCM
30	136	16.0	310	7	ADC31119	Adc31119 Human nov
31	135	15.8	82	4	AAM94589	Aam94589 Human rep
32	126	14.8	99	4	AAM95002	Aam95002 Human rep
33	126	14.8	99	4	ABB95710	Abb95710 Human tes
34	123.5	14.5	359	7	ADC31642	Adc31642 Human nov
35	98	11.5	1738	2	AAR50971	Aar50971 Norwalk v
36	98	11.5	1738	7	ADC72175	Adc72175 Norwalk v
37	97	11.4	260	3	AAG55557	Aag55557 Arabidops
38	97	11.4	274	3	AAG55556	Aag55556 Arabidops
39	95	11.2	180	3	AAG13988	Aag13988 Arabidops
40	95	11.2	259	3	AAG45886	Aag45886 Arabidops
41	85.5	10.0	212	6	ADA35397	Ada35397 Acinetoba
42	82.5	9.7	252	3	AAG55260	Aag55260 Arabidops
43	82	9.6	1181	4	ABG11187	Abg11187 Novel hum
44	78.5	9.2	380	4	ABB58658	Abb58658 Drosophil
45	77.5	9.1	197	5	ABB54252	Abb54252 Lactococc

ALIGNMENTS

RESULT 1
AAW85736
ID AAW85736 standard; protein; 164 AA.
XX
AC AAW85736;
XX
DT 01-OCT-1999 (first entry)
XX
DE Human analogue of human H-rev107 protein homologue.
XX
KW Transmembrane protein; transmembrane domain; translocation protein;
KW endoplasmic reticulum; sec61 alpha subunit; research; probe; diagnosis;
KW gene therapy; nutritional supplement; cytokine proliferation;
KW cell proliferation; cell differentiation; haematopoiesis regulation;
KW tissue growth; activin; inhibin; chemotactic activity; chemokinesis;
KW haemostasis; thrombolysis; receptor; ligand; anti-inflammatory;
KW tumour inhibition; antibody production.
XX
OS Homo sapiens.
XX
PN WO9918202-A2.
XX
PD 15-APR-1999.
XX
PF 05-OCT-1998; 98WO-JP004474.
XX
PR 08-OCT-1997; 97JP-00276269.
XX
PA (SAGA) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
XX
PI Kato S, Sekine S;
XX
DR WPI; 1999-277267/23.
DR N-PSDB; AAX08737, AAX08743.
XX
PT Human transmembrane proteins and nucleotide sequences.
XX
PS Claim 1; Page 70-71; 96pp; English.
XX
CC A novel method of cloning cDNAs from the human full-length cDNA bank is
CC described. The protein is identified by the presence of a hydrophobic N-
CC terminal secretory signal region, knowledge of the protein function is
CC not required, as in e.g. methods of expression cloning. The cDNAs encode
CC proteins comprising transmembrane domains and can be used as probes for
CC gene diagnosis and gene sources for gene therapy. Abnormalities of of
CC membrane proteins are associated with a number of hitherto-cryptogenic
CC diseases. e.g. cystic fibrosis. Polynucleotides and proteins of the

CC invention can be used in research; as nutritional supplements; for cell
CC proliferation; cell differentiation or cytokine activity or to induce the
CC production of cytokines; for immune stimulating or suppressing activity
CC i.e generation of antibodies; for haematopoiesis regulating activity; for
CC tissue growth activity; for activin/inhibin activity; for
CC chemotactic/chemokinetic activity; for haemostatic and thrombolytic
CC activity; for receptor/ligand activity; for anti-inflammatory activity or
CC for tumour inhibition activity
XX
SQ Sequence 164 AA;

Query Match 99.9%; Score 851; DB 2; Length 164;
Best Local Similarity 99.4%; Pred. No. 1.5e-89;
Matches 163; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60
Db 1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60

Qy 61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120
Db 61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120

Qy 121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFXIRRYQKKATA 164
Db 121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFAIRRYQKKATA 164

RESULT 2
AAW29736
ID AAW29736 standard; protein; 164 AA.
XX
AC AAW29736;
XX
DT 10-NOV-1998 (first entry)
XX
DE Human H-rev 107-like protein (HREVP).
XX
KW Human; H-rev 107-like protein; HREVP; prevention; diagnosis; cancer;
KW autoimmune disorder.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 154
FT /label= unknown
FT /note= "encoded by GNG"
XX
PN WO9836063-A1.
XX
PD 20-AUG-1998.
XX
PF 10-FEB-1998; 98WO-US002537.
XX
PR 14-FEB-1997; 97US-00801742.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Bandman O, Goli SK;
XX
DR WPI; 1998-457113/39.
DR N-PSDB; AAV47554.
XX
PT Nucleic acid encoding novel human H-rev 107-like protein - useful for
PT e.g. diagnosis, prevention and treatment of cancers or auto-immune
PT disorders.
XX
PS Claim 1; Page 41-42; 56pp; English.
XX
CC The present sequence represents human H-rev 107-like protein (HREVP). The
CC HREVP nucleic acid and protein can be used for the diagnosis, prevention
CC or treatment of cancer or autoimmune disorders
XX
SQ Sequence 164 AA;

Query Match 99.8%; Score 850; DB 2; Length 164;
Best Local Similarity 100.0%; Pred. No. 2e-89;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60
Db 1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60

Qy 61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120
Db 61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120

Qy 121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFXIRRYQKKATA 164
Db 121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFXIRRYQKKATA 164

RESULT 3
AAW60726
ID AAW60726 standard; protein; 164 AA.
XX
AC AAW60726;
XX
DT 02-SEP-1998 (first entry)
XX
DE Full length tazarotene inducible Gene-3 (TIG3) protein sequence.
XX
KW Tazarotene inducible gene-3; TIG3; identification; compound; treatment;
KW hyperproliferative skin disorder; psoriasis; acne; dysplasia; cancer.
XX
OS Homo sapiens.
XX
PN US5776687-A.
XX
PD 07-JUL-1998.
XX
PF 28-FEB-1997; 97US-00808303.
XX
PR 28-FEB-1997; 97US-00808303.
XX
PA (ALLR) ALLERGAN.
XX
PI Disepio D, Nagpal S, Chandraratna RA;
XX
DR WPI; 1998-398020/34.
DR N-PSDB; AAV36092.
XX
PT Identifying agents for treating hyper-proliferative skin diseases - from
PT ability to induce tazarotene inducible gene-3 mRNA in cultured skin
PT cells.
XX
PS Example 2; Col 29-32; 20pp; English.
XX
CC The present sequence is encoded by the full length cDNA seqeunce of
CC tazarotene inducible gene-3 (TIG3). The cDNA seqeunce was obtained by
CC alignment and combinantion of sequences AAV36088 and AAV36091. The
CC specification describes a method for identifying compounds suitable for
CC treating hyperproliferative disorders of the skin. The method comprises
CC analysing samples of untreated and treated cultures of skin cells for
CC presence of RNA from the TIG3 gene. If the treated sample has a higher
CC concentration of the RNA than the control, then the test compound is
CC identified as suitable for use as a treatment. The method is used to
CC identify compounds which are potentially useful for treating psoriasis,
CC acne or a wide range of dysplasias and cancers
XX
SQ Sequence 164 AA;

Query Match 98.6%; Score 840; DB 2; Length 164;
Best Local Similarity 98.2%; Pred. No. 2.8e-88;
Matches 161; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 10:51:49 ; Search time 21 Seconds
(without alignments)
751.210 Million cell updates/sec

Title: US-10-020-618-1
Perfect score: 852
Sequence: 1 MASPHQEPKPGDLIEIFRLG.....ILVVAGCSFXIRRYQKKATA 164

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	423.5	49.7	160	2	S42794	p18H-rev 107 prote
2	102	12.0	159	2	B82470	H-REV 107-related
3	92	10.8	1737	2	A37491	hypothetical helic
4	87.5	10.3	191	2	T50306	hypothetical prote
5	83.5	9.8	510	2	T44810	conserved hypothet
6	80	9.4	510	2	H69893	conserved hypothet
7	79.5	9.3	847	2	G95843	conserved hypothet
8	77.5	9.1	197	2	H86741	conserved hypothet
9	77	9.0	230	2	T34424	hypothetical prote
10	75.5	8.9	119	2	AC2351	hypothetical prote
11	75.5	8.9	338	2	T35517	probable secreted
12	75	8.8	400	2	S54642	glycine cleavage T
13	74.5	8.7	715	2	G86239	protein F20B24.6 [
14	74	8.7	453	2	T29493	hypothetical prote
15	73.5	8.6	671	2	AI0809	DNA ligase (NAD) (
16	72.5	8.5	397	2	AC2006	hypothetical prote
17	72.5	8.5	720	2	T47221	replication licens
18	72	8.5	281	2	S18245	xylF protein - Pse
19	72	8.5	363	1	S44734	probable protein-t
20	72	8.5	408	2	G88535	protein B0523.1 [i
21	72	8.5	749	2	B86403	probable mutator-l
22	72	8.5	895	2	T02597	Mutator-like trans
23	72	8.5	1127	2	E86386	probable mutator-l
24	71.5	8.4	252	2	T14964	phage lambda-relat
25	71.5	8.4	270	2	F64050	glucosamine-6-phos
26	71.5	8.4	693	2	G82618	pilus biogenesis p
27	71.5	8.4	964	2	E71460	probable outer mem
28	71.5	8.4	1115	2	T09403	integrin alpha cha
29	71.5	8.4	1115	2	T09433	integrin alpha cha
30	71	8.3	107	2	T27713	hypothetical prote
31	71	8.3	129	2	G87615	conserved hypothet
32	71	8.3	553	2	A82513	methyl-accepting c
33	71	8.3	556	2	T47552	hypothetical prote
34	71	8.3	669	2	JC5662	hepatoma-derived g
35	71	8.3	1286	1	S38058	hypothetical prote
36	70.5	8.3	293	2	S66356	mannose/glucose-bi
37	70.5	8.3	400	2	S32879	lipA protein - Nei
38	70.5	8.3	647	2	S61973	hypothetical prote
39	70.5	8.3	671	2	C91039	DNA ligase [import
40	70.5	8.3	671	2	F85883	DNA ligase [import
41	70.5	8.3	690	2	S28222	peroxidase (EC 1.1
42	70	8.2	621	2	E90253	formate hydrogenly
43	70	8.2	702	2	A96959	oxygen-sensitive r
44	70	8.2	5032	1	A35041	ryanodine receptor
45	70	8.2	5035	1	I46646	ryanodine receptor

ALIGNMENTS

RESULT 1
S42794
p18H-rev 107 protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 05-Nov-1999
C;Accession: S42794
R;Schaefer, R.
submitted to the EMBL Data Library, November 1993
A;Reference number: S42794
A;Accession: S42794
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-160 <SCH>
A;Cross-references: EMBL:X76453; NID:g433962; PIDN:CAA53991.1; PID:g433963

Query Match 49.7%; Score 423.5; DB 2; Length 160;
Best Local Similarity 52.5%; Pred. No. 3.6e-34;
Matches 83; Conservative 24; Mismatches 50; Indels 1; Gaps 1;

```
Qy          4 PHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEVKRE 63
              |  |||||||||||||  | |||:|:|||||||||||||| ||||::|: | |::| | ||:|
Db          2 PIPEPKPGDLIEIFRPMYSHWAIYVGDGYVIHLAPPSEIPGAGAASIMSALTDKAIVKKE 61

Qy         64 RLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQLRYG 123
                | || |   |:|||  | || | |:  ||  |:|:|||:: | : | |||||| :||||
Db         62 LLRDVAGKDKYQVNNKHDKEYTPLPLNKIIQRAEELVGQEVLYRLTSENCEHFVNELRYG 121

Qy        124 KSRCKQVEKAKVEVGVATALGILVVAGCSFXIRRYQKK 161
                 | || :  |:|   | :|:  :      : | :|:
Db        122 VPRSDQV-RDTVKVATVTGVGLAALGLIGVMLSRNKKQ 158
```

RESULT 2
B82470
H-REV 107-related protein VCA0346 [imported] - Vibrio cholerae (strain N16961 serogroup C
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: B82470
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: B82470
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-159 <HEI>
A;Cross-references: GB:AE004372; GB:AE003853; NID:g9657741; PIDN:AAF96254.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VCA0346

A;Map position: 2

Query Match 12.0%; Score 102; DB 2; Length 159;
Best Local Similarity 25.4%; Pred. No. 0.013;
Matches 43; Conservative 19; Mismatches 59; Indels 48; Gaps 9;

```
Qy          6 QEPKPGDLIEIFRLG-YEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEVKRER 64
              |    ||:: :   | |:||:| : |           |   : |   :   |: |
Db          4 QNLNAGDIV-VSNFGVYQHWSL-VSDALC----------EKGLPMLISATQRNGTVQEET 51

Qy         65 LEDVVGGCCYRVNNSLDHEY-----QPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFV-- 117
               : |  |           | |      |||  ::  |:  :|   |||:  |||||||
Db         52 WDVVTQG---------KHTYPAKVTYDRPVSEVLELARSQIG-TWKYSLTDRNCEHFVKW 101

Qy        118 --------TQLRYGKSRC----------KQVEKAKVEVGVATALGILVV 148
                     ||:  | :            :  |    :| | ||| | |
Db        102 ATGLKMSSTQVIAGATGAVLGASLVGLYSENPKFAKFLGGALALGSLAV 150
```

RESULT 3
A37491
hypothetical helicase/polymerase polyprotein - Southampton virus
N;Alternate names: orf1 protein
C;Species: Southampton virus
C;Date: 03-Mar-1994 #sequence_revision 25-Apr-1997 #text_change 17-Mar-2000
C;Accession: A37491
R;Lambden, P.R.; Caul, E.O.; Ashley, C.R.; Clarke, I.N.
Science 259, 516-519, 1993
A;Title: Sequence and genome organization of a human small round-structured (Norwalk-lik
A;Reference number: A37491; MUID:93142023; PMID:8380940
A;Contents: small round-structured virus, SRSV, Norwalk virus, Norwalk-like virus, serot
A;Accession: A37491
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: genomic RNA
A;Residues: 1-1737 <LAM>
A;Note: sequence extracted from NCBI backbone (NCBIP:123456)
C;Superfamily: rabbit calicivirus RNA-directed RNA polymerase

Query Match 10.8%; Score 92; DB 2; Length 1737;
Best Local Similarity 24.6%; Pred. No. 2;
Matches 29; Conservative 20; Mismatches 45; Indels 24; Gaps 5;

```
Qy          4 PHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEVKRE 63
              |  ||  ||:|| :     |:::||| |  : :  |          : ||       | |
Db        137 PASEPTIGDMIEFYEGHIYHYSIYIGQGKTVGVHSP--------QAAFS-------VARV 181

Qy         64 RLEDVVGGCCYRVNNSLDHEYQPRPV-EVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120
                :: :    :||      | |:|   :      ||:  :   |: :: ||  |  |:
Db        182 TIQPI--AAWWRVC------YIPQPKHRLSYDQLKELENEPWPYAAITNNCFEFCCQV 231
```

RESULT 4
T50306
hypothetical protein SPAPYUG7.06 [imported] - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C;Accession: T50306
R;McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Saunders, D.; Harris, D.
submitted to the EMBL Data Library, January 2000
A;Reference number: Z25059
A;Accession: T50306
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-191 <MCD>
A;Cross-references: EMBL:AL136521; PIDN:CAB66315.1; GSPDB:GN00066; SPDB:SPAPYUG7.06
A;Experimental source: strain 972h(-); clone plasmid pYUG7
C;Genetics:
A;Gene: SPDB:SPAPYUG7.06
A;Map position: 1

Query Match 10.3%; Score 87.5; DB 2; Length 191;
Best Local Similarity 31.0%; Pred. No. 0.43;
Matches 31; Conservative 17; Mismatches 33; Indels 19; Gaps 6;

```
Qy         24 WALYIGDGYVIHLAPPSEYP-GA----GSSSVFSVLSNSAEVKRERLEDVVGGCCYRVNN 78
              | | :|  :   :     ||  ||     ||: ||       | : |  ||     || :| :
Db         11 WTLGLGIYHTGLVLEGKEYAFGAHEIPGSTGVF------ATMPRPPLE----GCRWRCSI 60

Qy         79 SLDHEYQPRP--VEVIISSAKEMVGQKMKYSIVSRNCEHF 116
              :| :   |:|    ::|  ::|   |  : ||:: ||| ||
Db         61 ALPNCTLPKPDVDRILIRLSQEFTG--LSYSLLERNCNHF 98
```

RESULT 5
T44810
conserved hypothetical protein yngK [imported] - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 28-Jul-2000
C;Accession: T44810
R;Duitman, E.H.; Hamoen, L.W.; Rembold, M.; Venema, G.; Seitz, H.; Saenger, W.; Bernhard
Proc. Natl. Acad. Sci. U.S.A. 96, 13294-13299, 1999
A;Title: The mycosubtilin synthetase of Bacillus subtilis ATCC6633: A multifunctional hyl
A;Reference number: Z22848; MUID:20027541; PMID:10557314
A;Accession: T44810
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-510 <DUI>
A;Cross-references: EMBL:AF184956; NID:g6449053; PIDN:AAF08799.1; PID:g6449059
A;Experimental source: strain ATCC6633
C;Genetics:
A;Gene: yngK
C;Superfamily: conserved hypothetical protein yngK

Query Match 9.8%; Score 83.5; DB 2; Length 510;
Best Local Similarity 25.0%; Pred. No. 3.3;
Matches 29; Conservative 25; Mismatches 41; Indels 21; Gaps 6;

```
Qy         23 HWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSN--SAEVKRERLEDVVGGCCYRVNNSL 80
              ||   | :| : ::||   :    ::: : ||::  | |||   :   :|   |::||:
Db        294 HW---IQEGDIDYIAPQIYWSIGFNAAAYDVLADWWSNEVKNRPVHLYIGQAAYKINNNF 350

Qy         81 D------HEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHF------VTQLRYGK 124
              |        ||     :: ::   |:|   | :|:   |   |      :|:| | |
Db        351 DPPWSDPEEYVR---QITLNRQLELVKGSMHFSLKDLNKNPFGIKDRLITEL-YSK 402
```

RESULT 6
H69893
conserved hypothetical protein yngK - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 28-Jul-2000
C;Accession: H69893
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: H69893
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-510 <KUN>
A;Cross-references: GB:Z99113; GB:AL009126; NID:g2634090; PIDN:CAB13711.1; PID:g2634211
A;Experimental source: strain 168
C;Genetics:
A;Gene: yngK

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 10:48:44 ; Search time 18 Seconds
(without alignments)
474.417 Million cell updates/sec

Title: US-10-020-618-1
Perfect score: 852
Sequence: 1 MASPHQEPKPGDLIEIFRLG.....ILVVAGCSPXIRRYQKKATA 164

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	851	99.9	164	1	TIG3_HUMAN	Q9ul19 homo sapien
2	423.5	49.7	160	1	HRS3_RAT	P53817 rattus norv
3	418.5	49.1	162	1	HRS3_HUMAN	P53816 homo sapien
4	407.5	47.8	162	1	HRS2_HUMAN	Q9nww9 homo sapien
5	382	44.8	270	1	HRP5_MOUSE	Q9cpx5 mus musculu
6	380	44.6	279	1	HRP5_HUMAN	Q96kn8 homo sapien
7	350	41.1	168	1	HRS1_HUMAN	Q9hdd0 homo sapien
8	332.5	39.0	167	1	HRS1_MOUSE	Q9qzu4 mus musculu
9	92	10.8	1788	1	POLN_SOUV3	Q04544 southampton
10	77	9.0	458	1	GNFR_HUMAN	Q9ugk8 homo sapien
11	75	8.8	400	1	GCST_YEAST	P48015 saccharomyc
12	72.5	8.5	397	1	YG01_ANASP	P46015 anabaena sp
13	72.5	8.5	720	1	MCM7_XENLA	Q91876 xenopus lae
14	72	8.5	281	1	XYLF_PSEPU	P23106 pseudomonas
15	72	8.5	437	1	KI31_CAEEL	P34265 caenorhabdi
16	71.5	8.4	270	1	NAGB_HAEIN	P44538 haemophilus
17	71.5	8.4	624	1	MUTL_XANCP	Q8p8e4 xanthomonas
18	71.5	8.4	964	1	PMPE_CHLTR	O84877 chlamydia t
19	71.5	8.4	1115	1	ITA3_DROME	O44386 drosophila
20	71	8.3	1286	1	YKV5_YEAST	P28273 saccharomyc
21	70.5	8.3	293	1	LEC1_CLALU	Q39528 cladrastis
22	70.5	8.3	647	1	NOG1_YEAST	Q02892 saccharomyc
23	70.5	8.3	690	1	PERO_DROME	Q01603 drosophila
24	70	8.2	5035	1	RYR1_PIG	P16960 sus scrofa
25	70	8.2	5037	1	RYR1_RABIT	P11716 oryctolagus
26	70	8.2	5038	1	RYR1_HUMAN	P21817 homo sapien
27	69.5	8.2	377	1	TDP_DROME	Q24318 drosophila
28	69.5	8.2	505	1	PGKA_TRYBB	P08891 trypanosoma
29	69.5	8.2	3178	1	YS89_CAEEL	Q09624 caenorhabdi
30	69	8.1	320	1	DPA4_BPR69	O64301 bacteriopha
31	69	8.1	348	1	N33_HUMAN	Q13454 homo sapien
32	69	8.1	705	1	RNR_AQUAE	O67834 aquifex aeo
33	68.5	8.0	406	1	METK_METJA	Q58605 methanococc
34	68.5	8.0	632	1	S6AB_HUMAN	P48066 homo sapien
35	68.5	8.0	671	1	DNLJ_ECOLI	P15042 escherichia
36	68	8.0	162	1	CYTX_ONCVO	P22085 onchocerca
37	68	8.0	697	1	AD26_MOUSE	Q9r158 mus musculu
38	67.5	7.9	249	1	HYUE_PSESN	Q00924 pseudomonas
39	67.5	7.9	5327	1	MACP_MOUSE	Q9qxz0 mus musculu
40	67	7.9	772	1	CADJ_HUMAN	Q9h159 homo sapien
41	66.5	7.8	217	1	MMC1_METAC	P58867 methanosarc
42	66.5	7.8	248	1	KAD_GIALA	P49982 giardia lam
43	66.5	7.8	292	1	DHI1_SHEEP	P51975 ovis aries
44	66.5	7.8	359	1	MTD_FRAAN	Q9zrf1 fragaria an
45	66.5	7.8	444	1	HSLU_HAEIN	P43773 haemophilus

ALIGNMENTS

RESULT 1
TIG3_HUMAN
ID TIG3_HUMAN STANDARD; PRT; 164 AA.
AC Q9UL19; O95200;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Retinoic acid receptor responder protein 3 (Tazarotene-induced gene 3
DE protein) (RAR-responsive protein TIG3) (Retinoid inducible gene 1
DE protein).
GN RARRES3 OR TIG3 OR RIG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Jiang S.-Y., Huang S.-L., Shyu R.-Y., Yeh M.-Y.;
RT "Cloning and characterization of a novel retinoic acid inducible gene
RT RIG1 from human gastric cancer cells.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99061949; PubMed=9843971;
RA DiSepio D., Ghosn C., Eckert R.L., Deucher A., Robinson N., Duvic M.,
RA Chandraratna R.A.S., Nagpal S.;
RT "Identification and characterization of a retinoid-induced class II
RT tumor suppressor/growth regulatory gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:14811-14815(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Kato S.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";

```
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: May be a growth regulator that mediates some of the
CC       growth suppressive effects of retinoids.
CC   -!- TISSUE SPECIFICITY: Expressed in many tissues. Not expressed in
CC       heart, brain, and testis.
CC   -!- INDUCTION: By tazarotene.
CC   -!- SIMILARITY: Belongs to the H-rev107 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF092922; AAF02294.1; -.
DR   EMBL; AF060228; AAC84000.1; -.
DR   EMBL; AB030815; BAB08109.1; -.
DR   EMBL; BC009678; AAH09678.1; -.
DR   Genew; HGNC:9869; RARRES3.
DR   MIM; 605092; -.
DR   InterPro; IPR007053; NC.
DR   Pfam; PF04970; NC; 1.
FT   CONFLICT     63      63       E -> G (IN REF. 2).
FT   CONFLICT    118     118       T -> A (IN REF. 2).
SQ   SEQUENCE   164 AA;  18179 MW;  95625B8FD375FD38 CRC64;

  Query Match             99.9%;  Score 851;  DB 1;  Length 164;
  Best Local Similarity   99.4%;  Pred. No. 3.5e-77;
  Matches  163;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60

Qy         61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120

Qy        121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFXIRRYQKKATA 164
              |||||||||||||||||||||||||||||| ||||||||||
Db        121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFAIRRYQKKATA 164


RESULT 2
HRS3_RAT
ID   HRS3_RAT       STANDARD;      PRT;   160 AA.
AC   P53817;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   HRAS-like suppressor 3 (H-rev 107 protein).
GN   HRASLS3 OR HREV107 OR H-REV107.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=94119589; PubMed=8290259;
RA   Hajnal A., Klemenz R., Schaefer R.;
RT   "Subtraction cloning of H-rev107, a gene specifically expressed in
RT   H-ras resistant fibroblasts.";
RL   Oncogene 9:479-490(1994).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND ALSO IN CELL MEMBRANES.
CC   -!- SIMILARITY: Belongs to the H-rev107 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X76453; CAA53991.1; -.
DR   PIR; S42794; S42794.
DR   InterPro; IPR007053; NC.
DR   Pfam; PF04970; NC; 1.
SQ   SEQUENCE   160 AA;  17778 MW;  A0B32D6FA0BD12E6 CRC64;

  Query Match             49.7%;  Score 423.5;  DB 1;  Length 160;
  Best Local Similarity   52.5%;  Pred. No. 7.6e-35;
  Matches   83;  Conservative   24;  Mismatches   50;  Indels    1;  Gaps    1;

Qy          4 PHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEVKRE 63
              |  |||||||||||  | |||:|:|||||||||||||| ||||::|: | |:: | ||:|
Db          2 PIPEPKPGDLIEIFRPMYSHWAIYVGDGYVIHLAPPSEIPGAGAASIMSALTDKAIVKKE 61

Qy         64 RLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQLRYG 123
               | || |   |:|||  | || | |:  ||  |:|:|||:: | : | |||||| :||||
Db         62 LLRDVAGKDKYQVNNKHDKEYTPLPLNKIIQRAEELVGQEVLYRLTSENCEHFVNELRYG 121

Qy        124 KSRCKQVEKAKVEVGVATALGILVVAGCSFXIRRYQKK 161
               |  ||  :  |:|   | :|:  :      : | :|:
Db        122 VPRSDQV-RDTVKVATVTGVGLAALGLIGVMLSRNKKQ 158


RESULT 3
HRS3_HUMAN
ID   HRS3_HUMAN     STANDARD;      PRT;   162 AA.
AC   P53816; Q9HDD1;
DT   01-OCT-1996 (Rel. 34, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   HRAS-like suppressor 3 (H-rev 107 protein homolog) (HREV107-3).
GN   HRASLS3 OR HREV107.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=98442971; PubMed=9771974;
RA   Husmann K., Sers C., Fietze E., Mincheva A., Lichter P., Schafer R.;
RT   "Transcriptional and translational downregulation of H-REV107, a class
RT   II tumour suppressor gene located on human chromosome 11q11-12.";
RL   Oncogene 17:1305-1312(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kato S.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Siegrist S., Feral C., Chami M., Solhonne B., Rajpert-De Meyts E.,
RA   Guellaen G., Bulle F.;
RT   "Homo sapiens testis HREV107-3 mRNA.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2004, 10:50:09 ; Search time 42 Seconds
(without alignments)
1232.023 Million cell updates/sec

Title: US-10-020-618-1
Perfect score: 852
Sequence: 1 MASPHQEPKPGDLIEIFRLG.....ILVVAGCSFXIRRYQKKATA 164

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	411.5	48.3	162	11	Q8R3U1	Q8r3u1 mus musculu
2	390.5	45.8	168	11	Q8BWF7	Q8bwf7 mus musculu
3	380	44.6	269	4	Q8NE88	Q8ne88 homo sapien
4	345	40.5	168	4	Q86WS9	Q86ws9 homo sapien
5	258.5	30.3	235	5	Q8T778	Q8t778 branchiosto
6	150.5	17.7	231	11	Q9JI61	Q9ji61 rattus norv
7	146.5	17.2	230	4	O95237	O95237 homo sapien
8	146.5	17.2	230	4	Q8N716	Q8n716 homo sapien
9	146.5	17.2	230	6	Q9BGL2	Q9bgl2 bos taurus
10	146.5	17.2	231	11	Q9JI60	Q9ji60 mus musculu
11	136	16.0	310	4	Q96KN1	Q96kn1 homo sapien
12	127.5	15.0	292	4	Q8TAM5	Q8tam5 homo sapien
13	125	14.7	240	16	Q8EGZ0	Q8egz0 shewanella
14	123.5	14.5	292	4	Q96KN4	Q96kn4 homo sapien
15	123.5	14.5	292	4	Q86UZ2	Q86uz2 homo sapien
16	122.5	14.4	292	11	Q9D650	Q9d650 mus musculu
17	121.5	14.3	292	11	Q99K34	Q99k34 mus musculu
18	115.5	13.6	2134	12	Q9YLS4	Q9yls4 avian encep
19	108	12.7	227	16	Q7V994	Q7v994 prochloroco
20	102	12.0	159	16	Q9KMJ7	Q9kmj7 vibrio chol
21	98	11.5	1789	12	Q83883	Q83883 norwalk vir
22	97	11.4	260	10	Q8LBP8	Q8lbp8 arabidopsis
23	96	11.3	199	16	Q8EDA8	Q8eda8 shewanella
24	95	11.2	259	10	Q93V51	Q93v51 arabidopsis
25	94.5	11.1	242	10	Q9FFF0	Q9fff0 arabidopsis
26	93.5	11.0	283	10	Q9FFE7	Q9ffe7 arabidopsis
27	92.5	10.9	212	16	Q7UA38	Q7ua38 synechococc
28	92	10.8	1788	12	Q8JVV6	Q8jvv6 norwalk-lik
29	91	10.7	1787	12	Q8JXJ1	Q8jxj1 norwalk-lik
30	89	10.4	1785	12	Q9DU47	Q9du47 chiba virus
31	87.5	10.3	201	3	Q8X1T0	Q8x1t0 schizosacch
32	86.5	10.2	398	16	Q9KY68	Q9ky68 streptomyce
33	85	10.0	1699	12	Q8JXI6	Q8jxi6 norwalk-lik
34	83.5	9.8	337	16	Q82AJ4	Q82aj4 streptomyce
35	83.5	9.8	510	2	Q9R9I7	Q9r9i7 bacillus su
36	83	9.7	358	16	Q82HP1	Q82hp1 streptomyce
37	82.5	9.7	252	10	Q8L9G9	Q8l9g9 arabidopsis
38	82.5	9.7	252	10	Q9M872	Q9m872 arabidopsis
39	82	9.6	347	16	Q9KY71	Q9ky71 streptomyce
40	81.5	9.6	2256	12	Q8JV19	Q8jv19 ljungan vir
41	81	9.5	1784	12	Q9YS15	Q9ys15 norwalk vir
42	80	9.4	510	16	O35015	O35015 bacillus su
43	79.5	9.3	847	16	Q92XE7	Q92xe7 rhizobium m
44	79.5	9.3	2253	12	Q8JV20	Q8jv20 ljungan vir
45	79	9.3	1697	12	Q8JJL0	Q8jjl0 norwalk-lik

ALIGNMENTS

RESULT 1
Q8R3U1
ID Q8R3U1 PRELIMINARY; PRT; 162 AA.
AC Q8R3U1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to HRAS-like suppressor 3.
GN HRASLS3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC024581; AAH24581.1; -.
DR MGD; MGI:2179715; Hrasls3.
DR InterPro; IPR007053; NC.
DR Pfam; PF04970; NC; 1.
SQ SEQUENCE 162 AA; 17858 MW; EB180D8CF00AB2AB CRC64;

Query Match 48.3%; Score 411.5; DB 11; Length 162;
Best Local Similarity 51.9%; Pred. No. 6.8e-35;
Matches 84; Conservative 21; Mismatches 56; Indels 1; Gaps 1;

```
Qy          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60
              | :|  |||||||||||||  | |||:|:||||||||||||||  |||::|: | |:: | |
Db          1 MLAPIPEPKPGDLIEIFRPMYRHWAIYVGDGYVIHLAPPSEVAGAGAASIMSALTDKAIV 60

Qy         61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120
              |:|  |  | |   |:|||  | || | |:  ||  |: :|||:: | : | ||||| :|
Db         61 KKELLCHVAGKDKYQVNNKHDEEYTPLPLSKIIQRAERLVGQEVLYRLTSENCEHFVNEL 120

Qy        121 RYGKSRCKQVEKAKVEVGVA-TALGILVVAGCSFXIRRYQKK 161
              |||  |  ||  |  ||:|  |  | : |      : ||:
Db        121 RYGVPRSDQVRDAVKAVGIAGVGLAALGLVGVMLSRNKKQKQ 162
```

```
RESULT 2
Q8BWF7
ID   Q8BWF7      PRELIMINARY;      PRT;   168 AA.
AC   Q8BWF7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   H-REV 107 protein homolog.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK052657; BAC35085.1; -.
DR   InterPro; IPR007053; NC.
DR   Pfam; PF04970; NC; 1.
SQ   SEQUENCE   168 AA;  18932 MW;  73F410B77F832256 CRC64;

  Query Match             45.8%;  Score 390.5;  DB 11;  Length 168;
  Best Local Similarity   53.3%;  Pred. No. 1.1e-32;
  Matches   81;  Conservative   20;  Mismatches   48;  Indels    3;  Gaps    2;

Qy          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60
              | :|  |||||||||||  | |||:|:||||||||||||  |||::|: | |:: | |
Db          1 MLAPIPEPKPGDLIEIFRPMYRHWAIYVGDGYVIHLAPPSEIAGAGAASIMSALTDKAIV 60

Qy         61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120
              |:| |  | |   |:|||  | || | |:  ||  |: :|||:: | : | |||||| :|
Db         61 KKELLCHVAGKDKYQVNNKHDEEYTPLPLSKIIQRAERLVGQEVLYRLTSENCEHFVNEL 120

Qy        121 RYGKSRCKQV--EKAKVEVGVATALGILVVAG 150
              |||  |  ||      : :||  :   | |||
Db        121 RYGVPRSDQVCHRPCTLTLGV-RSFSQLPVAG 151


RESULT 3
Q8NE88
ID   Q8NE88      PRELIMINARY;      PRT;   269 AA.
AC   Q8NE88;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Similar to H-rev107-like protein 5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC034222; AAH34222.1; -.
DR   InterPro; IPR007053; NC.
DR   Pfam; PF04970; NC; 1.
SQ   SEQUENCE   269 AA;  29257 MW;  2018B203C2BFC8FB CRC64;

  Query Match             44.6%;  Score 380;  DB 4;  Length 269;
  Best Local Similarity   54.2%;  Pred. No. 2.6e-31;
  Matches   77;  Conservative   21;  Mismatches   40;  Indels    4;  Gaps    2;

Qy          8 PKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEVKRERLED 67
              |:|||||||||:||||||:|: |  |:|||||||     |  |: |: || | ||  ||||
Db        120 PRPGDLIEIFRIGYEHWAIYVEDDCVVHLAPPSEEFEVG--SITSIFSNRAVVKYSRLED 177

Qy         68 VVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQLRYGKSRC 127
              |: || ::||| ||  | | ||: ||   |:|| : ::||::  ||||||  ||||  |
Db        178 VLHGCSWKVNNKLDGTYLPLPVDKIIQRTKKMVNKIVQYSLIEGNCEHFVNGLRYGVPRS 237

Qy        128 KQVEKAKVEVGVATALGILVVA 149
              :||| | :|   | | | :: |
Db        238 QQVEHALME--GAKAAGAVISA 257


RESULT 4
Q86WS9
ID   Q86WS9      PRELIMINARY;      PRT;   168 AA.
AC   Q86WS9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   HRAS-like suppressor.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC048095; AAH48095.1; -.
DR   InterPro; IPR007053; NC.
DR   Pfam; PF04970; NC; 1.
SQ   SEQUENCE   168 AA;  18751 MW;  1184423034D8B39B CRC64;

  Query Match             40.5%;  Score 345;  DB 4;  Length 168;
  Best Local Similarity   49.0%;  Pred. No. 6.4e-28;
  Matches   71;  Conservative   24;  Mismatches   42;  Indels    8;  Gaps    2;

Qy          8 PKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEVKRERLED 67
              | |||||:|| ||:|||||:||||||::||     | |  :|  || |:  || : |:|
Db         15 PCPGDLIEVFRPGYQHWALYLGDGYVINIAPVDGIP-ASFTSAKSVFSSKALVKMQLLKD 73

Qy         68 VVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQLRYGKSRC 127
              |||    ||:||  |  | | ||| ||  :: ::||:: |:::  :||||||  ||||:
Db         74 VVGNDTYRINNKYDETYPPLPVEEIIKRSEFVIGQEVAYNLLVNDCEHFVTLLRYGEGVS 133

Qy        128 KQ-------VEKAKVEVGVATALGI 145
              :|       ||     ||| : ||:
Db        134 EQANRAISTVEFVTAAVGVFSFLGL 158


RESULT 5
Q8T778
ID   Q8T778      PRELIMINARY;      PRT;   235 AA.
AC   Q8T778;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Branchiostoma floridae (Florida lancelet) (Amphioxus).
OC   Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC   Branchiostoma.
OX   NCBI_TaxID=7739;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=11967531;
RA   Abi-Rached L., Gilles A., Shiina T., Pontarotti P., Inoko H.;
RT   "Evidence of en bloc duplication in vertebrate genomes.";
RL   Nat. Genet. 0:0-0(2002).
DR   EMBL; AF391288; AAM18866.1; -.
DR   InterPro; IPR007053; NC.
DR   Pfam; PF04970; NC; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   235 AA;  27172 MW;  D88AC7DE1E1F5FC8 CRC64;
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 10:55:44 ; Search time 22 Seconds
(without alignments)
384.848 Million cell updates/sec

Title: US-10-020-618-1
Perfect score: 852
Sequence: 1 MASPHQEPKPGDLIEIFRLG.....ILVVAGCSFXIRRYQKKATA 164

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	850	99.8	164	4	US-08-801-742-1	Sequence 1, Appli
2	840	98.6	164	1	US-08-808-303-8	Sequence 8, Appli
3	840	98.6	164	3	US-08-996-533-8	Sequence 8, Appli
4	839	98.5	164	1	US-08-808-303-12	Sequence 12, Appl
5	839	98.5	164	3	US-08-996-533-12	Sequence 12, Appl
6	623	73.1	124	1	US-08-808-303-3	Sequence 3, Appli
7	623	73.1	124	3	US-08-996-533-3	Sequence 3, Appli
8	423.5	49.7	160	4	US-08-801-742-4	Sequence 4, Appli
9	417.5	49.0	162	4	US-08-801-742-3	Sequence 3, Appli
10	98	11.5	1738	4	US-08-486-049-2	Sequence 2, Appli
11	85.5	10.0	212	4	US-09-328-352-6684	Sequence 6684, Ap
12	72.5	8.5	462	4	US-09-328-352-7128	Sequence 7128, Ap
13	71.5	8.4	674	4	US-09-489-039A-9158	Sequence 9158, Ap
14	71.5	8.4	964	4	US-09-556-877-177	Sequence 177, App
15	71.5	8.4	964	4	US-09-620-412C-177	Sequence 177, App
16	71.5	8.4	964	4	US-09-598-419-177	Sequence 177, App
17	71.5	8.4	977	4	US-09-556-877-191	Sequence 191, App
18	71.5	8.4	977	4	US-09-620-412C-191	Sequence 191, App
19	71.5	8.4	977	4	US-09-598-419-191	Sequence 191, App
20	71	8.3	235	2	US-08-760-745-1	Sequence 1, Appli
21	70	8.2	287	3	US-09-105-697-7	Sequence 7, Appli
22	70	8.2	830	1	US-07-977-434-6	Sequence 6, Appli
23	70	8.2	830	1	US-08-458-819-6	Sequence 6, Appli
24	70	8.2	830	5	PCT-US91-07035-6	Sequence 6, Appli
25	69	8.1	347	3	US-08-445-515-58	Sequence 58, Appl
26	69	8.1	348	3	US-08-445-515-56	Sequence 56, Appl
27	68.5	8.0	599	4	US-09-252-991A-19543	Sequence 19543, A
28	68.5	8.0	632	1	US-08-295-814E-10	Sequence 10, Appl
29	68.5	8.0	632	3	US-09-343-361-10	Sequence 10, Appl
30	68.5	8.0	632	5	PCT-US93-01959-10	Sequence 10, Appl
31	68.5	8.0	671	4	US-09-708-426-11	Sequence 11, Appl
32	68.5	8.0	916	4	US-09-688-078-9	Sequence 9, Appli
33	68	8.0	162	4	US-08-849-303-25	Sequence 25, Appl
34	68	8.0	234	4	US-09-800-729-150	Sequence 150, App
35	68	8.0	452	4	US-09-489-039A-11013	Sequence 11013, A
36	68	8.0	526	1	US-08-298-426-4	Sequence 4, Appli
37	67.5	7.9	179	4	US-09-489-039A-9458	Sequence 9458, Ap
38	67	7.9	287	3	US-09-105-697-8	Sequence 8, Appli
39	67	7.9	379	4	US-09-489-039A-11814	Sequence 11814, A
40	66.5	7.8	160	4	US-09-107-532A-7249	Sequence 7249, Ap
41	66.5	7.8	230	4	US-09-328-352-7052	Sequence 7052, Ap
42	66.5	7.8	501	2	US-08-408-095-31	Sequence 31, Appl
43	66.5	7.8	969	2	US-08-548-159-1	Sequence 1, Appli
44	66.5	7.8	986	2	US-08-548-159-3	Sequence 3, Appli
45	66.5	7.8	2391	2	US-08-446-855A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-801-742-1
; Sequence 1, Application US/08801742
; Patent No. 6359123
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL H-REV107-LIKE
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,742
; FILING DATE: Filed Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0200 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 164 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-801-742-1

Query Match 99.8%; Score 850; DB 4; Length 164;
Best Local Similarity 100.0%; Pred. No. 2e-94;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60

```
Db          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60

Qy         61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120

Qy        121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFXIRRYQKKATA 164
              |||||||||||||||||||||||||||||||||||||||||||||
Db        121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFXIRRYQKKATA 164
```

RESULT 2
US-08-808-303-8
; Sequence 8, Application US/08808303
; Patent No. 5776687
; GENERAL INFORMATION:
; APPLICANT: Nagpal, Sunil
; APPLICANT: DiSepio, Daniel
; APPLICANT: Chandraratna, Roshantha
; TITLE OF INVENTION: RETINOID INDUCED GENE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,303
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: ALRGN.062A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 164 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-808-303-8

Query Match 98.6%; Score 840; DB 1; Length 164;
Best Local Similarity 98.2%; Pred. No. 3.2e-93;
Matches 161; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSGVFSVLSNSAEV 60

Qy         61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KRGRLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120

Qy        121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFXIRRYQKKATA 164
              ||||||||||||||||||||||||||||| ||||||||||
Db        121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFAIRRYQKKATA 164
```

RESULT 3
US-08-996-533-8
; Sequence 8, Application US/08996533
; Patent No. 6294657
; GENERAL INFORMATION:
; APPLICANT: Nagpal, Sunil
; APPLICANT: DiSepio, Daniel
; APPLICANT: Chandraratna, Roshantha
; TITLE OF INVENTION: RETINOID INDUCED GENE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/996,533
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/808,303
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: ALRGN.062A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 164 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-996-533-8

Query Match 98.6%; Score 840; DB 3; Length 164;
Best Local Similarity 98.2%; Pred. No. 3.2e-93;
Matches 161; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSGVFSVLSNSAEV 60

Qy         61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KRGRLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120

Qy        121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFXIRRYQKKATA 164
              ||||||||||||||||||||||||||||| ||||||||||
Db        121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFAIRRYQKKATA 164
```

RESULT 4
US-08-808-303-12
; Sequence 12, Application US/08808303
; Patent No. 5776687
; GENERAL INFORMATION:
; APPLICANT: Nagpal, Sunil
; APPLICANT: DiSepio, Daniel

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: March 31, 2004, 10:58:15 ; Search time 40 Seconds
(without alignments)
1072.813 Million cell updates/sec

Title: US-10-020-618-1
Perfect score: 852
Sequence: 1 MASPHQEPKPGDLIEIFRLG.....ILVVAGCSFXIRRYQKKATA 164

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters: 1065169

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	851	99.9	164	12	US-10-616-942-2	Sequence 2, Appli
2	851	99.9	164	12	US-10-616-942-16	Sequence 16, Appl
3	850	99.8	164	13	US-10-020-618-1	Sequence 1, Appli
4	839	98.5	164	14	US-10-171-311-198	Sequence 198, App
5	423.5	49.7	160	13	US-10-020-618-4	Sequence 4, Appli
6	417.5	49.0	162	12	US-10-616-942-26	Sequence 26, Appl
7	417.5	49.0	162	12	US-10-616-942-27	Sequence 27, Appl
8	417.5	49.0	162	13	US-10-020-618-3	Sequence 3, Appli
9	415.5	48.8	169	9	US-09-925-300-1341	Sequence 1341, Ap
10	350	41.1	168	12	US-10-616-942-1	Sequence 1, Appli
11	350	41.1	168	12	US-10-616-942-14	Sequence 14, Appl
12	345	40.5	237	10	US-09-764-891-3842	Sequence 3842, Ap
13	135	15.8	82	10	US-09-764-891-3247	Sequence 3247, Ap
14	126	14.8	99	10	US-09-764-891-3660	Sequence 3660, Ap
15	99.5	11.7	296	12	US-10-425-114-64719	Sequence 64719, A
16	98	11.5	1738	14	US-10-314-739-2	Sequence 2, Appli
17	95	11.2	311	12	US-10-425-114-64839	Sequence 64839, A
18	93	10.9	304	12	US-10-425-114-70402	Sequence 70402, A
19	91	10.7	329	12	US-10-425-114-58957	Sequence 58957, A
20	90.5	10.6	259	12	US-10-424-599-142890	Sequence 142890,
21	90	10.6	201	12	US-10-424-599-168222	Sequence 168222,
22	87.5	10.3	260	12	US-10-424-599-181930	Sequence 181930,
23	86.5	10.2	194	12	US-10-425-114-58343	Sequence 58343, A
24	83.5	9.8	337	14	US-10-156-761-13590	Sequence 13590, A
25	83	9.7	358	14	US-10-156-761-11004	Sequence 11004, A
26	80.5	9.4	393	12	US-10-424-599-262457	Sequence 262457,
27	80	9.4	410	15	US-10-369-493-9975	Sequence 9975, Ap
28	79.5	9.3	305	12	US-10-425-114-53077	Sequence 53077, A
29	79.5	9.3	319	12	US-10-425-114-71733	Sequence 71733, A
30	79	9.3	141	12	US-10-424-599-181931	Sequence 181931,
31	79	9.3	278	12	US-10-425-114-61459	Sequence 61459, A
32	76.5	9.0	783	12	US-10-282-122A-61551	Sequence 61551, A
33	76.5	9.0	857	15	US-10-369-493-10874	Sequence 10874, A
34	76	8.9	363	14	US-10-156-761-11542	Sequence 11542, A
35	75.5	8.9	173	9	US-09-738-626-6015	Sequence 6015, Ap
36	75.5	8.9	4881	14	US-10-156-761-8481	Sequence 8481, Ap
37	74.5	8.7	392	12	US-10-425-114-64608	Sequence 64608, A
38	74	8.7	453	15	US-10-369-493-6714	Sequence 6714, Ap
39	73.5	8.6	398	14	US-10-156-761-11001	Sequence 11001, A
40	73.5	8.6	500	14	US-10-081-051-40	Sequence 40, Appl
41	73.5	8.6	669	12	US-10-282-122A-73417	Sequence 73417, A
42	73.5	8.6	671	9	US-09-815-242-13864	Sequence 13864, A
43	73.5	8.6	671	12	US-10-282-122A-75582	Sequence 75582, A
44	72.5	8.5	432	15	US-10-369-493-3436	Sequence 3436, Ap
45	72.5	8.5	836	12	US-10-425-114-51748	Sequence 51748, A

ALIGNMENTS

RESULT 1
US-10-616-942-2
; Sequence 2, Application US/10616942
; Publication No. US20040048339A1
; GENERAL INFORMATION:
; APPLICANT: Kato, Seishi
; APPLICANT: Sekine, Shingo
; TITLE OF INVENTION: HUMAN PROTEINS HAVING TRANSMEMBRANE DOMAINS AND
; TITLE OF INVENTION: CDNAS ENCODING THESE PROTEINS
; FILE REFERENCE: GIN-6710CPUS
; CURRENT APPLICATION NUMBER: US/10/616,942
; CURRENT FILING DATE: 2003-07-11
; PRIOR APPLICATION NUMBER: US/09/529,100
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: JP 0276269
; PRIOR FILING DATE: 1997-10-08
; PRIOR APPLICATION NUMBER: PCT/JP98/04474
; PRIOR FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-942-2

Query Match 99.9%; Score 851; DB 12; Length 164;
Best Local Similarity 99.4%; Pred. No. 1.3e-85;
Matches 163; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60

Qy         61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120
```

```
Qy        121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFXIRRYQKKATA 164
              |||||||||||||||||||||||||||||||| ||||||||||
Db        121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFAIRRYQKKATA 164


RESULT 2
US-10-616-942-16
; Sequence 16, Application US/10616942
; Publication No. US20040048339A1
; GENERAL INFORMATION:
;  APPLICANT: Kato, Seishi
;  APPLICANT:  Sekine, Shingo
;  TITLE OF INVENTION: HUMAN PROTEINS HAVING TRANSMEMBRANE DOMAINS AND
;  TITLE OF INVENTION:  CDNAS ENCODING THESE PROTEINS
;  FILE REFERENCE: GIN-6710CPUS
;  CURRENT APPLICATION NUMBER: US/10/616,942
;  CURRENT FILING DATE:  2003-07-11
;  PRIOR APPLICATION NUMBER: US/09/529,100
;  PRIOR FILING DATE: 2000-08-21
;  PRIOR APPLICATION NUMBER: JP 0276269
;  PRIOR FILING DATE: 1997-10-08
;  PRIOR APPLICATION NUMBER: PCT/JP98/04474
;  PRIOR FILING DATE: 1998-10-05
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
;   LENGTH: 164
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-616-942-16

  Query Match             99.9%;  Score 851;  DB 12;  Length 164;
  Best Local Similarity   99.4%;  Pred. No. 1.3e-85;
  Matches  163;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60

Qy         61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120

Qy        121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFXIRRYQKKATA 164
              |||||||||||||||||||||||||||||||| ||||||||||
Db        121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFAIRRYQKKATA 164


RESULT 3
US-10-020-618-1
; Sequence 1, Application US/10020618
; Publication No. US20020156256A1
;   GENERAL INFORMATION:
;        APPLICANT: Bandman, Olga
;                   Goli, Surya K.
;        TITLE OF INVENTION: A NOVEL H-REV107-LIKE
;                            PROTEIN
;        NUMBER OF SEQUENCES: 4
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Incyte Pharmaceuticals, Inc.
;             STREET: 3174 Porter Drive
;             CITY: Palo Alto
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 94304
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/020,618
;             FILING DATE: 06-Dec-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/801,742
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Billings, Lucy J.
;             REGISTRATION NUMBER: 36,749
;             REFERENCE/DOCKET NUMBER: PF-0200 US
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 415-855-0555
;             TELEFAX: 415-845-4166
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 164 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-020-618-1

  Query Match             99.8%;  Score 850;  DB 13;  Length 164;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-85;
  Matches  164;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MASPHQEPKPGDLIEIFRLGYEHWALYIGDGYVIHLAPPSEYPGAGSSSVFSVLSNSAEV 60

Qy         61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KRERLEDVVGGCCYRVNNSLDHEYQPRPVEVIISSAKEMVGQKMKYSIVSRNCEHFVTQL 120

Qy        121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFXIRRYQKKATA 164
              |||||||||||||||||||||||||||||||| ||||||||||
Db        121 RYGKSRCKQVEKAKVEVGVATALGILVVAGCSFAIRRYQKKATA 164


RESULT 4
US-10-171-311-198
; Sequence 198, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel, Robert
;  APPLICANT:  Chen, Yan
;  APPLICANT:  Zhao, Xumei
;  APPLICANT:  Monahan, John
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Glatt, Karen
;  APPLICANT:  Gannavarapu, Manjula
;  APPLICANT:  Hoersh, Sebastian
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
;  TITLE OF INVENTION:  OF CERVICAL CANCER
;  FILE REFERENCE: MRI-035
;  CURRENT APPLICATION NUMBER: US/10/171,311
;  CURRENT FILING DATE:  2002-06-12
;  PRIOR APPLICATION NUMBER: US 60/298,159
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,155
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/335,936
;  PRIOR FILING DATE: 2001-11-14
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 198
;   LENGTH: 164
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-171-311-198

  Query Match             98.5%;  Score 839;  DB 14;  Length 164;
  Best Local Similarity   98.2%;  Pred. No. 2.7e-84;
  Matches  161;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```